DOCUMENT RESUME

ED 330 342 TM 019 104

AUTHOR Barnhart, Amber J.; And Others
TITLE A Comparison of Second- and Fourth-Year Medical Students on a Standardized-Patient Examination of Clinical Competence: A Construct Validity Study.
PUB DATE 91
NOTE 13p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Clinical Diagnosis; Comparative Analysis; *Competence; *Construct Validity; Effect Size; Evaluation Methods; Graduate Medical Education; Higher Education; Medical Evaluation; *Medical Students; Patients; *Test Validity
IDENTIFIERS Patient Management; *Southern Illinois University School of Medicine

ABSTRACT

The construct validity of the standardized-patient (SP) examination used at Southern Illinois University (Springfield) School of Medicine was assessed by comparing 66 second-year and 70 fourth-year medical students on 5 SP cases. The results show sizable differences between the groups. The usefulness of passing rates the effect-size measures as a means of enhancing the typically weak evidence for validity provided by group-differences studies of construct validity is demonstrated. Results obtained through these approaches show that, as would be expected, the clerkships are having a considerable effect on clinical competence and that the examination is sensitive, as a valid measure of clinical competence should be, to these changes in the clinical-competence construct. A table summarizes passing rates and group means. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED350342

A Comparison of Second- and Fourth-Year Medical Students on a Standardized-Patient Examination of Clinical Competence: A Construct Validity Study

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
AMBER J. BARNHART
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Amber J. Barnhart, M.D.

Michelle L. Marcy, M.S.

Jerry A. Colliver, Ph.D.

Steven J. Verhulst, Ph.D.

Southern Illinois University

School of Medicine

Springfield, Illinois

Author responsible for Correspondence and Reprint Requests:

Amber J. Barnhart, M.D.

Department of Family Practice

Southern Illinois University

School of Medicine

Springfield, Illinois 62708

TM019104

BEST COPY AVAILABLE

ABSTRACT

The construct validity of the standardized-patient (SP) examination used at Southern Illinois University School of Medicine was assessed by comparing second-year and fourth-year medical students on five SP cases. The results showed sizeable differences between the groups. The paper demonstrates the usefulness of passing rates and effect-size measures as a means of enhancing the typically weak evidence for validity provided by group-differences studies of construct validity. The results obtained with these approaches show that the clerkships are having a considerable effect on clinical competence as would be expected and that the examination is sensitive to these changes in the clinical-competence construct as a valid measure of clinical competence should.

The use of standardized patients (SPs) for assessing clinical competence has increased rapidly in recent years.[1] In response to this increase, a large body of research on the psychometric properties of SP-based assessments has emerged.[2] Much of this research has dealt with reliability. These studies have demonstrated the reproducibility of examination scores and pass-fail decisions[2] and have shown that the reproducibility is generally unaffected by potentially disruptive factors.[3,4] A few studies have focused on the construct validity of these assessments by showing that groups of examinees at more advanced levels of training perform better on the SP cases than do examinees at earlier levels.[5-9] The rationale for this group-differences approach to construct validity is that the validity of a measure of a given construct is supported if the measure is sensitive to and reflects differences among groups thought to differ on the construct.[10]

Three of the group-differences studies compared residents at different points in their residency training programs.[5,6,7] Another study compared first- and second-year residents with third-year medical students,[8] and another compared fifth- and sixth-year medical students and residents.[9] In general, the results of these studies supported the construct validity of the SP-based assessments of clinical competence, by showing that the clinical competence construct was increasing with additional clinical training as would be expected and as should be reflected by a valid measure of clinical competence.

At Southern Illinois University (SIU) School of Medicine, a performance-based examination of clinical competence that uses SP cases is given to all senior medical students upon completion of their clinical clerkships.[11,12] Students are expected to pass the examination to fulfill a part of their graduation requirements. The purpose of this Post-Clerkship Examination is to determine if students have mastered the clinical competencies expected of

students upon completion of their clerkships as defined by School of Medicine objectives.

The present study was conducted to assess the construct validity of the SIU SP-based Post-Clerkship Examination. To accomplish this, five SP cases that were used as a part of the Post-Clerkship Examination administered to seniors in the class of 1990 were selected for administration to second-year students in the class of 1992 upon completion of their Introduction to Clinical Medicine course. Thus, the present study provides a comparison of second-year and fourth-year medical students, groups tested before and after their first major clinical experience, respectively, in the clinical clerkships. In addition to comparing the means of the second- and fourth-year students which is the typical analytic approach that has been used in group-differences studies of construct validity, in this study the passing rates of the two groups were compared and strength of effect measures were computed, to provide a quantitative indication of the impact of the clerkship training on clinical competence and the sensitivity of the Post-Clerkship Examination to differences between the second- and fourth-year student groups.

Methods

The Examination. A thorough discussion of the SIU Post-Clerkship Examination including details of the development, administration and scoring is presented elsewhere and should be consulted for a full description of the examination.[11,12] In brief, the examination is a performance-based examination that uses about 18 forty minute SP cases (20 minutes for the student-SP encounter and another 20 minutes immediately following the encounter for students to answer written questions about the case). Cases for the examination are chosen by the faculty Post-Clerkship Examination committee and represent the most frequently encountered patient problems as well as the

most important patient problems that students are expected to evaluate and manage competently. Competencies to be assessed by each case are determined by the faculty committee. Faculty physicians provide the patient cases and, with an educator, develop the instruments for collecting student performance data. These data consist of checklists completed by SPs who record actions performed by students on history and physical, and written responses by students following the patient encounters to questions concerning findings, tentative diagnostic conclusions, and plans for treatment and management. For each case, a passing level is established by the case author and reviewed by the faculty committee. This Case Pass Level reflects the standards of performance expected of senior medical students upon completion of their clerkship rotations as expressed by minimal scores on the clinical competencies being assessed by the case. For this study, a student was said to have passed the full examination, consisting of all five SP cases, if his or her total examination score (i.e., the mean of all of the student's case scores) exceeded the mean of the Case Pass Levels.

Analytic Methods. For the present study, five SP cases were randomly selected from 13 of the 18 cases that comprised the Post-Clerkship Examination given to the class of 1990 (n = 70) in their fourth year of medical school (tested in October, 1989). Five cases that emphasized management, pediatrics, or OB/GYN problems were excluded from the random selection process because they were not thought appropriate for second-year students. The five cases selected were then administered to second-year students in the class of 1992 (n = 66) upon completion of their Introduction to Clinical Medicine course (tested in May, 1990). The fourth-year students were tested after completion of all clerkship rotations and the second-year students were tested before their clerkships were started. The means and passing rates of the second-year

and fourth-year students were compared on each of the five selected cases with t tests and χ^2 tests, respectively. Differences between the means in pooled standard deviation units and odds ratios were computed, in order to assess the strength of the effect of the clinical clerkship experience and the sensitivity of the examination to the effect of this experience.

Results and Discussion

The results of the statistical tests provide good support for the construct validity of the SP assessment. (See Table 1.) For total scores, which are averages across all five cases, the mean of the fourth-year group (74.59) was significantly higher than that of the second-year group (62.43) ($p = .0001$), and the passing rate of the fourth-year group (70%) was significantly higher than that of second-year group (3%) ($p = .001$). For all five cases, the means and passing rates were significantly higher for the fourth-year students than for the second-year students ($p < .05$).

However, as pointed out by writers from Cronbach and Meehl[10] to van der Vleuten and Swanson[2], the demonstration of significant group differences constitutes at best weak evidence for validity. The problem, in part, is that it is not clear how much the groups should differ. If an expected difference in the amount of a construct (e.g., clinical competence) possessed by members of the different groups were specified, an empirical confirmation of the specified difference with the measuring instrument would provide strong support for the construct validity of the measure. To address this problem, passing rates and strength of effect measures were computed, in addition to the tests of significance on the means. Although it is not clear how big the group differences should be, the effect size measures indicate how big they are, so that it is possible to judge, at least intuitively, whether they seem

to be of reasonable magnitude given our conceptual understanding of the construct.

The passing rates, in particular, provided especially informative evidence regarding the magnitude of group-difference issue. For total scores, which are averages across all five cases, the difference was considerable, with 70% of the fourth-year students passing and only 3% of the second-year students passing. Clearly, the clinical experience in the clerkships was a virtual necessity for passing the five-case, clinical-competence examination. The magnitude of the difference between the passing rates (3% versus 70%) would seem to be consistent with our intuitive expectation of the magnitude of the difference in clinical competence between second-year and fourth-year students, resulting from their first major clinical experience. Similarly, the odds ratio for this difference was 78.20, indicating that the odds of a fourth-year student passing the examination is 78 times greater than the odds of a second-year student passing. Again, the magnitude of effect given by the odds ratio would seem to be consistent with our intuitive expectation, indicating a sizeable effect. Even for means, the difference between the total score means was 2.19 standard deviations, indicating that average performance of fourth-year students was over 2 standard deviations higher than that of second-year students.

Conclusion

The significance of the study is that the results show sizeable differences between the second- and fourth-year groups (i.e., groups assessed before and after their clinical clerkships, respectively) and thus provide good support for the construct validity of the Post-Clerkship Examination as a measure of clinical competence. Moreover, the paper demonstrates the usefulness of passing rates and effect-size measures as a means of enhancing

the typically weak evidence for validity provided by group-differences studies. For example, the comparison of passing rates provides a more intuitively meaningful estimate of the magnitude of the effect of the clerkship rotations on clinical competence. The use of odds ratios with the passing rates provides another intuitively meaningful way of estimating the effect size. Even the expression of mean differences in standard deviation units adds to the meaning of usual significance test assessment of group-differences. The results obtained with these approaches show that the clerkships are having a considerable effect on clinical competence as would be expected and that the Post-Clerkship Examination is sensitive to these changes in the clinical-competence construct as a valid measure of clinical competence should.

Acknowledgement

The authors wish to acknowledge Raymond A. Gensinger, Jr., M.D. for his help with the scoring of the SP cases in the ICM course.

References

1. Hart, I.R. More Developments in Assessing Clinical Competence. Montreal, Quebec: Can-Heal Publications Inc., 1991 (in press).

2. van der Vleuten C. P.M., Swanson, D.B. Assessment of clinical skills with standardized patients: State of the art. Teaching and Learning in Medicine: An International Journal 2(1990): 58-76.

3. Colliver, J.A., Morrison, L.J., Markwell, S.J., Verhulst, S.J., Steward, D.E., Dawson-Saunders, B., Barrows, H.S. Three studies of the effect of multiple standardized patients on intercase reliability of five standardized-patient examination. Teaching and Learning in Medicine: An International Journal 2(1990):237-245.

4. Colliver, J.A., Steward, D.E., Markwell, S.J., Marcy, M.L. Effect of repeated simulations by standardized patients on intercase reliability. Teaching and Learning in Medicine: An International Journal 3(1991):15-19.

5. Stillman, P., Swanson, D., Smee, S., et al. Assessing clinical skills of residents with standardized patients. Annals of Internal Medicine 105(1986):762-771.

6. Stillman, P., Rutala, P., Nicholson, G., Sabers, S., Stillman, A. Measurement of clinical competence of residents using patient instructors. In Proceedings of the 21st Research in Medical Education

Conference (pp.111-6). Washington, DC: Association of American Medical Colleges, 1982.

7. Robb, K., Rothman, A. The assessment of clinical skills in general internal medicine residents -- Comparison of the objective structured clinical examination to a conventional oral examination. Annals of the Royal College of Physicians and Surgeons of Canada 18(1985):235-238.

8. Petrusa, E., Blackwell, T., Parcel, S., Saydjari, C. Psychometric properties of the objective clinical exam as an instrument for final evaluation. In I. Hart, R. Harden, H. Walton (Eds.), Newer developments in assessing clinical competence (pp. 181-91). Montreal: Heal, 1986.

9. Newble, D., Hoare, J., Elmslie, R. The validity and reliability of a new examination of the clinical competence of medical students. Medical Education 17(1981):165-71.

10. Cronbach, L.J., Meehl, P.E. Construct validity in psychological tests. Psychological Bulletin 52(1955) 281-302.

11. Barrows, H.A., Williams, R.G., Moy, R.H. A comprehensive performance-based assessment of fourth-year students clinical skills. Journal of Medical Education 62(1987):802-809.

12. Williams, R.G., Barrows, H.S., Vu, N.V., Verhulst, S.J., Colliver, J.A., Marcy, M.L., Steward, D.E. Direct, standardized assessment of clinical competence. Medical Education 21(1987):482-489.

Table 1. Means (± standard deviations) and passing rates with p values and strength of effect measures (d for means and OR for passing rates) for second-year (n = 70) and fourth-year (n = 66) students on five SP cases. d is difference between means in pooled standard deviation units and OR is odds ratio.

	Second-year Students	Fourth-year Students	p	Strength of effect
Means:				d
Case 1	67.79 (±9.28)	73.26 (±9.69)	.0010	.58
Case 2	57.36 (±9.92)	70.58 (±10.67)	.0001	1.29
Case 3	65.47 (±13.97)	71.39 (±13.89)	.0145	.42
Case 4	60.63 (±11.52)	84.27 (±9.58)	.0001	2.23
Case 5	60.93 (±9.90)	73.45 (±10.77)	.0001	1.21
Total	62.43 (±5.86)	74.59 (±5.21)	.0001	2.19
Passing Rates:				OR
Case 1	41%	68%	.002	3.03
Case 2	37%	80%	.001	6.90
Case 3	23%	44%	.009	2.65
Case 4	14%	89%	.001	50.57
Case 5	1%	38%	.001	42.07
Total	3%	70%	.001	78.20